

Magazyny danych – stan obecny i kierunki rozwoju

Robert WREMBEL, Zbyszko KRÓLIKOWSKI, Mikołaj MORZY

Politechnika Poznańska, Instytut Informatyki
ul. Piotrowo 3a, 60-965 Poznań

Otrzymano 24 lutego 2000 roku

Streszczenie. Magazyny danych są bardzo dużymi bazami danych, w których gromadzi się dane pochodzące z wielu heterogenicznych źródeł, np. innych scentralizowanych lub rozproszonych baz relacyjnych, relacyjno–obiektywnych, obiektywnych oraz ze źródeł innych niż bazy danych, takich, jak arkusze kalkulacyjne, pliki XML, zasoby WWW. Potrzeba gromadzenia takich danych jest szczególnie widoczna w sektorze handlowym. Analiza informacji o rynku, popycie i sprzedaży danych towarów, trendach i anomaliach jest jednym z etapów podejmowania strategicznych decyzji w przedsiębiorstwie. Analizę przeprowadza się zwykle za pomocą aplikacji OLAP wykorzystujących bardzo złożone zapytania i operujących na ogromnych ilościach danych, rzędu setek czy tysięcy gigabajtów. Kluczowym problemem w magazynach danych jest więc efektywność przetwarzania informacji. W ostatnim czasie wiele uwagi poświęcono opracowaniu modeli danych (wielowymiarowy model danych), metod fizycznej implementacji modelu wielowymiarowego (relacyjny i wielowymiarowy OLAP), nowych algorytmów optymalizacji zapytań (indeksy bitmapowe). Artykuł ten jest poświęcony omówieniu procesu konstruowania magazynu danych, zagadnień związanych z pielęgnacją i odświeżaniem magazynu oraz kwestiom architektury magazynu danych. W artykule przedstawiono nowy model przetwarzania (OLAP), wielowymiarowy model danych, metody jego implementacji oraz opisano podstawy zagadnienia optymalizacji zapytań w środowisku magazynu danych.

Słowa kluczowe: magazyn danych, OLAP, OLTP, analiza danych, wielowymiarowy model danych, kostka danych, indeks bitmapowy